

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-06-07

Total records in dataset: 26,616,527

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	26,616,527	26,616,527
total_cells	319,398,324	79,849,581
missing_fields	65,440,400	882,419
missing_pct	20%	1%
complete_fields	253,957,924	78,967,162
complete_pct	80%	99%
unknown_fields	22,366,356	9,519,783
unknown_pct	7%	12%
suppressed_fields	100	100
suppressed_pct	0%	0%
available_fields	231,591,468	69,447,279
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	12	0.0%	648,767	2.4%
sex	18	0.0%	53,494	0.2%
age_group	70	0.0%	180,158	0.7%
records_with_any_field	70	0.0%	821,755	3.1%